



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A Workshop Series for Grid/Repository Integration

Citation for published version:

Aschenbrenner, A, Blanke, T, Chue Hong, NP, Ferguson, N & Hedges, M 2009, 'A Workshop Series for Grid/Repository Integration' D-Lib Magazine, vol. 15, no. 1/2. DOI: 10.1045/january2009-aschenbrenner

Digital Object Identifier (DOI):

[10.1045/january2009-aschenbrenner](https://doi.org/10.1045/january2009-aschenbrenner)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Published In:

D-Lib Magazine

Publisher Rights Statement:

Copyright policy: <http://www.dlib.org/about.html#access>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



CONFERENCE REPORT

D-Lib Magazine
January/February 2009

Volume 15 Number 1/2

ISSN 1082-9873

A Workshop Series for Grid/Repository Integration[Andreas Aschenbrenner](#)State and University Library, Goettingen
<aschenbrenner@sub.uni-goettingen.de>[Tobias Blanke](#)King's College London, Arts and Humanities e-Science Support Centre (AHeSSC)
<tobias.blanke@kcl.ac.uk>[Neil P Chue Hong](#)OMII UK
<N.ChueHong@omii.ac.uk>[Nicholas Ferguson](#)OGF.eeig, OGF Europe
<n.ferguson@trust-itservices.com>[Mark Hedges](#)King's College London, Centre of e-Research (CeRch)
<mark.hedges@kcl.ac.uk>

Researchers and information scientists are often forced to move back and forth between different digital environments in their daily work. Institutional or thematic repositories have become a prevalent mechanism to manage publications, and increasingly also to manage research outputs and primary data. Integration of the user's natural work environment with repositories is improving (albeit slowly), and repository-based research environments are emerging (e.g., eSciDoc¹). The natural habitat for many scientific users, however, is e-Infrastructure like the "grid". At the same time, these users are employing repositories – often home-grown systems – to store their research outputs and publications.

A workshop series supported by OGF-Europe,² DReSNeT³ and OMII⁴ set out to reduce this fragmentation and explore the interfaces between grid- and repository-based architectures. Past workshops include:

- 5 June 2008, Open Grid Forum (OGF) 23, Barcelona - <http://www.ogf.org/OGF23/>
- 4 November 2008, DReSNeT Kick-Off, Newcastle - <http://www.dresnet.net/>
- 1 December 2008, Digital Curation Conference (DCC) 2008, Edinburgh - <http://www.dcc.ac.uk/events/dcc-2008/programme/>

- o 12 December 2008, IEEE e-Science 2008, Indianapolis - <http://escience2008.iu.edu/workshops/digital/index.shtml>

Each of these four workshops was attended by between 50 and 100 participants from a variety of backgrounds. Despite a tangible terminology gap between repository managers and some (scientific) users, the commonalities between requirements and existing systems were astonishing.

High on the agenda of all participants is preservation. Funding organizations or legal frameworks often demand the preservation of primary data for 10 years and more, and another factor is the inherent value of the primary data, e.g., such as irreproducible climate data. As Flavia Donno analyzed in her presentation at OGF23, the scientific and the repository communities need to share with each other their experiences in preservation and work towards open standards such as OAIS.⁵

Participants at all four workshops emphasised various aspects of preservation. Two related discussions included:

1. The capture of provenance data with regard to both the documentation of the (organizational) context, as well as the technical audit trail, since data is created through a number of steps in a scientific workflow;
2. The preservation of the creation process and the discourse about scientific results, e.g., by integrating Wikis, Blogs, etc., directly into the trusted repository.⁶

Many of the measures used to ensure preservation are also conducive to good scientific practice in general. This includes preservation of the context of data and publications. In her talk at OGF23, Francoise Genova's demonstration was on the Strasbourg Astronomical Observatory, which links publications and the data upon which they are based. Reliable metadata are, of course, essential for this. Indeed, a recurring topic at the workshops was metadata in all its aspects, from semi-automatic metadata creation to ensuring metadata quality.

All these issues – preservation, good scientific practice, metadata and collection management – emphasize the tight integration of the repository with the user's work environment where the data is created. Reliable audit trails, metadata, etc. can only be created if the users' work environments and repositories connect seamlessly. Moreover, as Andrew Treloar (Australian National Data Service, ANDS⁷) put it in the discussion at IEEE e-Science 2008: "Repositories are just the plumbing". All the tedious bits of repository management need to be hidden from the user wherever possible. At best, metadata should not be user-created, it should just be there; systems should interface on-the-fly; and data should not be accessioned into the repository, but be born and sustained in the repository without the user noticing.

The federation of distinct repositories is obviously an important factor for achieving this. Users must not be required to ingest data multiple times, e.g., into their institutional repository and into a thematic repository too. This should be the goal regardless of whether in the future there will be only one virtual repository based on multiple distinct repository systems, or whether there will be buckets of federated repositories addressing specific needs (e.g., data safety for confidential medical records). Because of this, the OAI-ORE protocol⁸ has been received with high interest by the scientific community as it resembles protocols dubbed "grid" in many ways. OAI-ORE allows for virtualization of digital repositories just as e.g., the OGSA-DAI⁹ protocol virtualizes distributed databases.

Essentially, all kinds of infrastructure need to care – as David de Roure's eloquent analysis at IEEE e-Science 2008 shows – more about the roll-in of users rather than roll-out of services. Users are understandably highly sensitive about tools being "delivered" to them by somebody who might be less proficient in the area of the users' work. In this respect, the users don't want to be "served", rather they are looking for simple and pragmatic environments with which they can interface. Hence, better interweaving of existing infrastructures is vital to ensure user satisfaction and eventually the persistence of each of those infrastructures.

How do we get there?

Of course, repositories and the grid are not the only kinds of infrastructure available, but interweaving them is an essential step towards the simple and pragmatic plumbing the user is seeking. The repository community has for a long time acknowledged that other components besides repositories are needed to achieve a suitable information environment. Format registries (e.g., Pronom,¹⁰ GDFR¹¹) or a persistent identifier resolver (e.g., Pili¹²) are some of the components that were identified during the workshop series. Grid-based workflow engines, storage services, and other modules could interface with repositories in a very similar way.

All possible technological considerations are, of course, caught between the underlying trends of experimentation and standardization. We may need more experimentation to reach the desired level of integration. However, the time for pure experimentation is coming to an end for both repositories and the grid. Funding institutions were represented at the workshops, and they are looking for operational, trustworthy scientific repositories with many of the features described above. For example, the Australian National Data Service (ANDS) has recently been commissioned (Andrew Treloar, IEEE e-Science 2008); the call for DataNet by the NSF in the USA¹³ is being evaluated (Ed Seidel, IEEE e-Science 2008 conference); and the latest call for e-Infrastructure scientific repositories by the European Commission¹⁴ is just out (Krystyna Marek, OGF23).

The next pragmatic step is the upcoming workshop of the series of grid/repository integration workshops. It is the Open Grid Forum 25, which will be held in Catania, on March 2-6, 2009.¹⁵ "Infrastructure" may be of many kinds, but let's ensure that these infrastructures can interweave.

Notes

1. eSciDoc, <<http://www.escidoc.org/>>.
2. OGF Europe, <<http://www.ogfeurope.eu/>>.
3. DReSNe, <<http://www.dresnet.net/>>.
4. OMII-UK, <<http://www.omii.ac.uk/>>.
5. OAIS - Open Archival Information System (OAIS) Reference Model.
<http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html>.
6. cf. Andreas Hense's talk at IEEE e-Science 2008,
<<http://live.escience2008.iu.edu/calendar/Calendar.html>>.

7. Australian National Data Service, <<http://ands.org.au/>>.
8. Open Archives Initiative Object Reuse and Exchange (OAI-ORE), <<http://www.openarchives.org/ore/>>.
9. OGSA-DAI. <<http://www.ogsadai.org.uk/>>.
10. PRONOM, <<http://www.nationalarchives.gov.uk/pronom/>>.
11. Global Digital Format Registry, <<http://www.gdfr.info/>>.
12. PILIN, <<https://www.pilin.net.au/>>.
13. DataNet forum, <<http://network.nature.com/forums/datanet/587>>.
14. e-Infrastructure Call for Participation, <<http://cordis.europa.eu/fp7/ict/e-infrastructure/>>.
15. OGF25/EGEE User Forum, <<http://www.ogf.org/OGF25/>>.

Copyright © 2009 Andreas Aschenbrenner, Tobias Blanke, Neil P. Chue Hong, Nicholas Ferguson, and Mark Hedges

[Top](#) | [Contents](#)
[Search](#) | [Author Index](#) | [Title Index](#) | [Back Issues](#)
[Previous Article](#) | [Next Conference Report](#)
[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)

doi:10.1045/january2009-aschenbrenner